Cyberclio

Marin Dacos

Cyberclio.

Vers une Cyberinfrastructure au coeur de la discipline historique

Marin Dacos
Directeur du Centre pour l'édition électronique ouverte – Cleo.cnrs.fr
Marin.dacos@revues.org
http://blog.homo-numericus.net

Été 2009. À la stupéfaction générale, le libraire en ligne Amazon.com procède à la suppression unilatérale des versions électroniques de *1984* et de *La ferme des animaux* dans les Kindle[1] de leurs propriétaires. La fiction imaginée par George Orwell dans *1984* semble dépassée par la réalité. Les lecteurs dépossédés ainsi de leurs exemplaires étaient-ils sanctionnés pour avoir piratés les ouvrages ? Un bug avait-il touché le géant américain de la librairie en ligne ? Ni piratage, ni bug : c'est un problème juridique qui s'est produit, car Amazon avait commercialisé certains ouvrages de façon aventureuse. Ce n'est pas la première fois que les pionniers du numérique prennent des libertés avec le droit traditionnel. Mais c'est la première fois que les honnêtes clients d'Amazon, heureux propriétaires d'une liseuse, de modèle Kindle et de marque Amazon, se voient dépossédés à leur insu d'un ouvrage régulièrement acquis. Amazon a pénétré dans les Kindle de ses clients et effacé les fichiers qu'ils contenaient. Un tel épisode s'est reproduit en décembre 2010, pour des ouvrages érotiques vendus par Amazon, puis retirés de son catalogue, ainsi que des Kindle auxquels ils avaient été vendus. La société américaine a refusé de rembourser les ouvrages et son service commercial s'est même permis de répondre aux clients lésés que leurs lectures étaient immorales [2]. En retirant *1984*, la société de Jeff Bezos a également détruit l'ensemble des annotations personnelles ajoutées par les lecteurs. Le scandale produit par ce premier exemple d'autodafé numérique perpétré par un libraire a eu un retentissement mondial et le supermarché des livres a dû faire marche arrière. L'histoire ne dit pas si les annotations ont été récupérées, ou si elles ont été perdues à jamais. Si elles ont été perdues, c'est le signe de la fragilité des dispositifs numériques dont disposent aujourd'hui les intellectuels, au rang desquels on aura la faiblesse d'intégrer les historiens. Si elles ont pu être récupérées, c'est sans doute pire, car cela signifie qu'Amazon réalise des copies de secours des données personnelles de ses clients sur ses serveurs.

Alors que la troisième révolution industrielle, la révolution numérique, atteint l'objet-livre, cette anecdote emblématique interroge les chercheurs sur la façon dont ils doivent mener leurs recherches, leur travail d'écriture et leur métier d'enseignant. En effet, à l'heure de l'informatique dans les nuages (*cloud computing* [3]), se pose la question de la maîtrise de sa bibliothèque numérique par le chercheur. Mais le problème est bien plus large, car il touche aussi les sources de l'historien, sujet encore plus complexe et

---

[1] Le Kindle est une liseuse de livres électroniques fabriquée et vendue par Amazon.

[2] http://theselfpublishingrevolution.blogspot.com/2010/12/amazon-in-book-banning-business.html

[3] Il s'agit d'une méthode informatique déportant les calculs et les données sur le réseau, alors qu'on les exploitait traditionnellement sur un ordinateur local. Les traitements de texte en ligne (Google docs, Zoho,…) constituent un exemple typique du *cloud computing*.

stratégique, d'un point de vue heuristique. Élargie à l'ensemble des secteurs de l'activité scientifique, du séminaire permanent en ligne à la maîtrise de l'identité numérique du chercheur, des problèmes d'annotations aux questions d'interopérabilités entre les corpus, le problème a tendance à jeter la communauté scientifique dans un abîme de perplexité. Face à un environnement numérique foisonnant, instable, imprédictible et soumis aux forces du marché, comment l'historien *lambda* peut-il être un acteur de son devenir numérique, et non subir les vagues d'innovations et les contre-vagues d'obsolescence accélérée, de modes éphémères et de faillites industrielles ? La réponse tient en un seul mot : cyberinfrastructure.

1. Retour aux sources ?

Un retour sur la célèbre enquête des « TRA », initiée au début des années 1980, permet d'interroger la question d'une base historique collective créée avant le développement du Web, la généralisation de l'interopérabilité des données et la naissance du mouvement « Open Science Data »[4]. Voici comment les porteurs de cette enquête la décrivent : « Commencée dans les années 80 par Jacques Dupâquier, l'enquête "TRA" vise à analyser les dynamiques démographiques et sociales de la population française de 1803 à 1986, en se fondant sur la reconstitution d'un corpus extrêmement important et unique de généalogies patronymiques descendantes (personnes dont le patronyme commence par les lettres TRA). Le choix d'un corpus patronymique permet d'utiliser les tables décennales alphabétiques de naissances, mariages et décès tenues dans toutes les mairies à partir de 1803. Le fait d'utiliser une sélection patronymique écarte les descendances des femmes mariées. Cette contrainte est nécessaire pour permettre une collecte aussi exhaustive que possible mais elle est corrigée par l'introduction à chaque génération des conjointes des hommes TRA. Le choix des noms TRA fournit un corpus de personnes représentées dans tous les départements (quelles que soient les langues locales : alsacien, catalan...) qui, tout en n'étant pas trop important, permet d'avoir un échantillon représentatif au niveau national d'environ 3000 couples, sur la base d'un couple pour 10000 habitants par département au recensement de 1806."[5]

Comment cette base informatisée, financée par des fonds publics, est-elle mutualisée avec la communauté des historiens ? Le site du Laboratoire de démographie historique mentionne des contraintes légales et une volonté de dépôt : "Cette enquête qui utilise des données nominatives est soumise à des contraintes administratives strictes en particulier de la CNIL : non diffusion d'informations permettant d'identifier les personnes, ce qui limite le contenu des bases de données qui seront mises à disposition. De plus l'ensemble des documents devra être déposé aux AN [Archives nationales] après la fin de l'enquête". Daté de 2008, ce texte confirme que les données ne sont pas publiques, ou du moins accessibles aux historiens, 25 ans après le début de cette enquête d'envergure. Il est sans doute possible d'entrer en contact avec les responsables de la base et de leur demander un accès, mais aucune procédure explicite ne permet de

---

[4] La libération des données de la recherche est réclamée dès 2003 dans la Déclaration de Berlin. http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/ Ce mouvement a été relancé par le développement d'initiatives gouvernementales de libération massive de données publiques : data.gov et data.gov.uk

[5] http://www.ehess.fr/ldh/theme_TRA/Theme_TRA-Intro.htm

savoir dans quelles conditions cela pourrait avoir lieu. On aurait tort de jeter la pierre aux porteurs de l'enquête TRA. Tout d'abord, parce qu'on imagine la difficulté de mettre à disposition une base de données qui aurait traversé 25 ans de mutations technologiques sans investir beaucoup de temps dans la documentation de l'histoire technique d'une telle base, dans la maintenance des données et dans la mise au point d'interfaces de consultation ou d'extraction à distance. Mais aussi, et surtout, parce que l'institution ne valoriserait pas un tel effort, que ce soit à court terme ou à moyen terme. En effet, la mutualisation des données est en général considérée comme une naïveté par la communauté des historiens : pourquoi investir temps et argent dans des opérations considérées comme techniques pour permettre à ses collègues, et néanmoins concurrents, de mener des études à peu de frais, sur la base de données collectées laborieusement ? Sans incitation et sans reconnaissance collective, la mutualisation de données scientifiques reste le parent pauvre de l'activité historique, elle est soumise aux variations de la déontologie individuelle et reste conditionnée à la mise en place de relations individuelles entre le détenteur des données et le chercheur souhaitant l'exploiter. Bref, rien ne prédispose les données historiques numériques à la mutualisation. Elles font partie, dans le meilleur des cas, du « web invisible »[6]. Dans le pire des cas, elles prennent la forme de bases de données disponibles dans un format propriétaire et/ou obsolète, disponible dans un ou deux bureaux d'une institution, soumise au risque de vol, de dégradation, de démagnétisation, d'obsolescence ou, pire encore, d'oubli. Car les données sont généralement incompréhensibles sans le savoir, essentiellement oral, permettant de connaître les modalités pratiques qui ont présidé à leur élaboration. Usages, conventions, sous-entendus, raccourcis et petites habitudes constituent une grille de lecture aussi importante que la documentation du logiciel utilisé ou les articles préfigurant la construction de la base.

Alors que la déontologie des historiens concernant son rapport aux sources est très rigoureuse et exigente, elle n'a connu aucun *aggiornamento* numérique. Les archéologues sont quasiment les seuls à se tenir à une obligation de publication du résultat de leurs fouilles, en raison du caractère destructif de leur travail de collecte. Cela confère à leur discipline une originalité de laquelle il faut s'inspirer. Les archéologues ont mis en place un écosystème éditorial complet, permettant de publier les résultats des fouilles. Pour les autres périodes, il existe des initiatives intéressantes, mais rien d'aussi systématique que la *Carte archéologique de la Gaule*[7].

Discipline d'origine littéraire, l'Histoire reste très souvent une entreprise individuelle, s'appuyant sur la figure traditionnelle de l'auteur et de son génie, rarement sur une démarche collective. Or, l'historien produit son savoir sur la base d'un matériau essentiellement inédit et original. Cela le transforme en ensemble quasiment exclusif, en

---

6 Le web invisible est constitué des contenus en ligne qui échappent aux moteurs de recherche.

7 La *Carte Archéologique de la Gaule* est une collection de l'Académie des Inscriptions et Belles-Lettres lancée en 1931, relancée en 1988, coéditée (depuis 1992) avec la Sous-Direction de l'Archéologie (Direction de l'Architecture et du Patrimoine), le Ministère de la Recherche, la Maison des Sciences de l'Homme. Cette collection est chargée de recenser, d'étudier et de publier, département par département, l'ensemble des découvertes archéologiques de la France de l'âge du Fer au début du Moyen Âge (c'est-à-dire de 800 av. J.-C. à 800 après J.-C.).

raison des difficultés d'accès aux différentes sources. Seul le savoir qui en découle est rendu public. La plupart du temps, les matériaux sur lesquels s'appuie le discours historique sont cités par extraits, mais rarement exposés publiquement dans leur totalité. Or, l'historien s'appuie sur les éléments les plus significatifs d'un corpus archivistique, éléments qui confortent la thèse, et a tendance à moins mettre en valeur les parasites qui ne "collent" pas avec elle. On devine la puissance heuristique que pourrait avoir la mise en commun d'au moins une partie des matériaux inédits extraits de chaque nouvelle enquête historique, que ce soit dans des archives publiques ou privées, imprimées ou manuscrites, textuelles, iconiques ou sérielles. La forme de cette mise en commun reste à débattre au sein de la profession, car il n'est pas envisageable de transformer l'historien en moine copiste. En revanche, lorsqu'il formalise son matériau en construisant son *corpus*, cette formalisation devrait devenir, à terme, le patrimoine commun de la profession, de l'université, voire, osons le mot, de l'humanité.

En effet, la généralisation du numérique et des réseaux constitue une opportunité historique pour recomposer le métier d'historien autour de la notion de travail collaboratif, susceptible de sortir le chercheur de son splendide isolement. À l'heure où les tenants du libre accès à la littérature scientifique s'interrogent sur une obligation de dépôt en ligne des publications des chercheurs (« *open access mandate* »[8]), on peut s'interroger sur l'extension de ce type d'obligation ou, au moins, d'incitation aux sources qui alimentent la démonstration de l'historien. Les modalités restent à définir. Le dépôt en ligne d'une base de données n'est pas le seul moyen : il est également possible d'adjoindre à un article des compléments électroniques qui appuient la démonstration et qui seraient susceptibles d'être réutilisés par d'autres. C'est la stratégie mise en œuvre par les sociologues de la revue *Sociologie*, fondée par Serge Paugam, qui ont décidé de la doter d'un supplément intitulé « Sociologie 2.0 ». Ce supplément, dirigé par Pierre Mercklé, se définit comme suit : « dans cette rubrique, la revue publie, en supplément électronique de chaque numéro de la version papier, un article scientifique original distingué par la mobilisation de procédés faisant un usage avancé de dispositifs numériques innovants d'argumentation, d'administration de la preuve et de documentation de la recherche (techniques d'enrichissement textuel, documents audiovisuels, représentations graphiques animées…) »[9]. Tout un programme ! Car, si le gain d'une telle démarche semble aller de soi, nombreuses sont les interrogations sur ce qui pourrait apparaître, pour beaucoup, comme un mirage technologique.

2. Le mirage technologique ?

Les annales des rapports entre histoire et informatique sont jalonnées de projets avortés, de données perdues et de formats inexploitables. Les cartes perforées des années 1970, les disquettes 5"¼ des années 1980, les bases de données propriétaires des années 1990, les erreurs 404[10] et les défaçages[11] des années 2000 sont restés dans

8 On appelle « mandat » les obligations de dépôt des résultats de la recherche mises en place par les institutions (Universités, Agences) à destination de leurs agents ou des programmes de recherches qu'elles financent. Lire, notamment Stevan Harnad, Les Carr, Alma Swan, Arthur Sale, Hélène Bosc, « Maximizing and Measuring Research Impact Through University and Research-Funder Open-Access Self-Archiving Mandates », 2009. http://archivesic.ccsd.cnrs.fr/sic_00493057/en/ Voir également Willinsky, John. 2006. *The access principle : the case for open access to research and scholarship*. Cambridge Mass, 2006, MIT Press.

9 http://sociologie.revues.org/155

10 Erreur provoquée par l'appel à une page web qui n'existe plus à cette adresse.

les mémoires et n'encouragent guère à investir la sphère numérique. En Sciences humaines et sociales, comme ailleurs, la « loi de Murphy » guette les entreprises. La loi de Murphy, qui veut qu'une catastrophe finit toujours par arriver, c'est-à-dire que la tartine a une propension à tomber du côté beurré, est très souvent utilisée par les informaticiens pour conjurer le mauvais sort avec humour.

La pauvreté de l'information est également une menace réelle, puisque le chercheur construit son corpus pour un usage très précis, dans lequel l'implicite règne en maître. En dehors de leur contexte initial, les bases de données risquent d'être inutilisables, incompréhensibles, voire introuvables. Ce dernier syndrome est bien connu des bibliothécaires, qui savent qu'un livre mal rangé dans une bibliothèque d'un million de livres est un livre perdu, ainsi que des spécialistes de l'information numérique, qui savent qu'une donnée mal décrite est une donnée perdue, dès lors qu'on la mutualise, au sein de milliers d'autres corpus. On pourrait appeler cela le syndrome de l'archipel, tellement dispersé que sa cartographie devient impossible. Par ailleurs, les documents sont initialement inertes et ne se lient pas entre eux. L'interopérabilité à l'intérieur du corpus, mais aussi entre les corpus, est donc quasiment nulle, ce qui produit un « effet tunnel », isolant les documents. Le syndrome du tunnel est, en général, renforcé par les processus de privatisation de données, que ce soit à l'initiative du chercheur isolé ou de l'organisme, public ou privé, qui prend en charge le corpus. Sans parler des tentations spéculatives, la mise en place des conditions d'accès aux corpus peut être motivée par des contraintes juridiques, techniques ou commerciales. Dans tous les cas, elle freine l'interopérabilité des corpus et réduit leur potentiel heuristique.

La réponse à l'ensemble de ces problèmes est la mise en place d'une cyberinfrastructure au service de l'histoire. Le nom a de quoi faire peur, puisqu'il semble promettre un « machin » technologique piloté par la technique, dans lequel l'historien aurait une place secondaire et sur lequel il aurait peu de prise. Il pourrait, en outre, annoncer un rupture profonde dans le métier d'historien, réalisant la prophétie d'Emmanuel Le Roy Ladurie selon laquelle l'historien sera programmeur ou ne sera pas[12]… Le projet d'une cyberinfrastructure pour historien signe-t-elle la fin d'une façon incarnée de faire l'histoire ? Cela surviendra si la cyberinfrastructure est conçue sans les principaux intéressés.

3. Cyberinfrastructure

Le rapport de l'ACLS sur les cyberinfrastuctures en sciences humaines et sociales définit les cyberinfrastructures comme « une couche d'information, d'expertise, de standards, de principes, d'outils et de services qui sont largement partagés entre les communautés de recherches, mais développés pour des besoins universitaires spécifiques : une cyberinfrastructure est plus spécifique que le réseau lui-même, mais est également plus générale qu'un outil ou une ressource développés pour un projet ou même pour une discipline en particulier » [13]. La réussite d'une cyberinfrastructure dépendra notamment

---

11 Remplacement de la page d'accueil d'un site par une page composée par ceux qui l'ont piratée. Habituellement, la page est remplacée par une tête de mort et par la signature du pirate.

12 Référence Ladurie, *Le territoire de l'historien*, 1973.

13 Cf. la note de Pierre Mounier au sujet de ce rapport http://blog.homo-

de sa capacité à se positionner au bon niveau d'intervention. Pour cela, je propose de distinguer programmes de recherches, plateformes et très grands équipements. L'ensemble de ces trois niveaux, harmonieusement articulés, constitue une cyberinfrastructure.

1) les *programmes de recherches* peuvent couvrir des équipes régionales, nationales ou même internationales, avoir une grande ampleur problématique, géographique ou chronologique, et mobiliser de nombreux chercheurs ainsi que de nombreuses ressources. Ils portent les problématiques et les innovations scientifiques. Financés en général par projet, ils ne disposent pas d'infrastructures numériques leur permettant de pérenniser leurs méthodes et leurs *corpus* sur le long terme (les laboratoires de recherche n'ont pas cette vocation). Ils publient leurs résultats au cours du programme, et une fois celui-ci achevé.

2) les *plateformes* sont des centres spécialisés en Humanités numériques qui mènent des missions de long terme à forte dimension technologique. À cette échelle, un effort de généricité est nécessaire pour transposer dans la longue durée des projets dont les modalités sont initialement pensées comme des prototypes uniques et spécifiques. Cet effort de généricité doit parvenir à respecter l'intégrité du questionnement scientifique de chaque projet, tout en se préoccupant :

- de réduction drastique des idiomes, grâce à l'adoption de normes internationales,
- de factorisation de l'ensemble des éléments technologiques mutualisables,
- de diffusion la plus large possible, dans le respect d'une politique d'accès contrôlée à chaque fois que cela s'impose (données nominatives, données confidentielles, ...),
- de conservation et d'accès à long terme.

Les plateformes ont donc des compétences fortes en ingénierie (modélisation des données et des processus, développement, documentation) et s'appuient sur des dispositifs technologiques puissants, afin de stabiliser l'effort d'innovation scientifique issu des projets de recherches.

3) les *infrastructures*, qui assurent le financement des plateformes et arbitrent sur les priorités stratégiques. Leur rôle est également d'assurer l'interconnexion et la mise en cohérence de l'ensemble des dispositifs des plateformes. Ils s'assurent que l'évolution des plateformes est en phase avec l'évolution des enjeux mis en évidence par la communauté scientifique. Ces grands équipements portent sur les quatre dimensions majeures de la recherche en sciences humaines et sociales :

- *l'accès*, la stabilisation et la mise en relation des données numériques entre elles,
- *les données* sur lesquelles s'appuient les chercheurs (archives historiques, enquêtes orales, statistiques diverses, données archéologiques...) *mais aussi les méthodes et outils* qui permettent d'en extraire des découvertes scientifiques,
- *l'édition* des résultats de la recherche (livres, revues, archives ouvertes),
- *la vie des communautés scientifiques* (débat scientifique, identité numérique).

Ils constituent les formes contemporaines des infrastructures traditionnelles des sciences humaines et sociales : bibliothèques, archives, presses universitaires, universités, centres de recherches.

Les travaux de construction des Cyberinfrastructures sont en cours. On citera le projet

---

numericus.net/article130.html et le rapport lui-même : *Our Cultural Commonwealth, The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*, American Council of Learned Societies, New York, 2006, 43 pages.

Bamboo aux Etats-Unis[14], Joint Information Systems Committee (JISC)[15] au Royaume-Uni, les initiatives françaises (PROGEDO, CORPUS, BSN [16]) dont le Très Grand Équipement Adonis[17] et son moteur de recherche[18], les projets européens DARIAH[19] et CLARIN[20]. D'autres initiatives jouent un rôle très structurant sans être des cyberinfrastructures, comme le W3C[21] ou le Consortium TEI (TEI-C)[22]. Ces organismes jouent un rôle de définition de normes et de guides de bonnes pratiques qui jouent un rôle essentiel.

L'ÉTAT DE L'ART

L'ensemble a vocation à obéir à l'état de l'art, tant d'un point de vue scientifique que d'un point de vue technologique. Cette formulation, simple et rassurante, cache une réalité complexe et évolutive, sur laquelle la stabilisation d'un savoir positif est difficile. Il ne suffit pas, en effet, de lister l'ensemble des technologies et des normes auxquelles il faut obéir pour être conformes à l'état de l'art. Le « lancer » de sigles technologiques (« sigles dropping » sur le modèle du « name dropping ») est, en effet, un sport qui demande peu d'efforts et qui a vocation à rassurer un auditoire réputé ignorant dans le domaine technologique.

En revanche, on connaît quelques-uns des grands principes qu'il faut respecter et des technologies à appliquer. C'est l'objectif que se donne le Manifeste des Digital Humanities, rédigé par une centaine de spécialistes lors de THATCamp Paris en mai 2010[23]. Selon ce texte, la communauté des Humanités numériques se structure autour de grands principes : interopérabilité, ouverture et documentation des formats, échange de bonnes pratiques. Ce domaine ne pourra se développer et se structurer sans la mise en place d'une communauté de professionnels spécialisés, affectés à temps plein à ce type de mission, et dont les compétences seront mises à jour en permanence. Dans un contexte de forte concurrence salariale dans le secteur informatique, cette ambition n'est pas la plus simple à assumer sur la durée. Les Humanités numériques sont en train de se doter de guides de bonnes pratiques. On consultera, pour la France, les guides de bonnes pratiques rédigés par le TGE Adonis[24], pour l'Angleterre ceux du JISC[25]. On trouvera ci-dessous deux illustrations de la nature et de la complexité des objets

---

14 http://projectbamboo.uchicago.edu/

15 http://www.jisc.ac.uk/

16 http://www.roadmaptgi.fr/Documents/TGIRs%20en%20STIC%20et%20SHS.pdf Voir également nos propositions : Jean-Paul Caverni, Marin Dacos, *Construire les Digital humanities en France. Des Cyber-infrastructures pour les Sciences humaines et sociales,* Rapport remis à la la Commission des présidents d'université (CPU), 2009, 15 pages.

17 http://www.tge-adonis.fr/

18 http://rechercheisidore.fr/

19 http://www.dariah.eu/

20 http://www.clarin.eu

21 http://www.w3.org/

22 http://www.tei-c.org

23 Version française : http://tcp.hypotheses.org/318 et version anglaise : http://tcp.hypotheses.org/411

24 http://www.tge-adonis.fr/ressources/guides

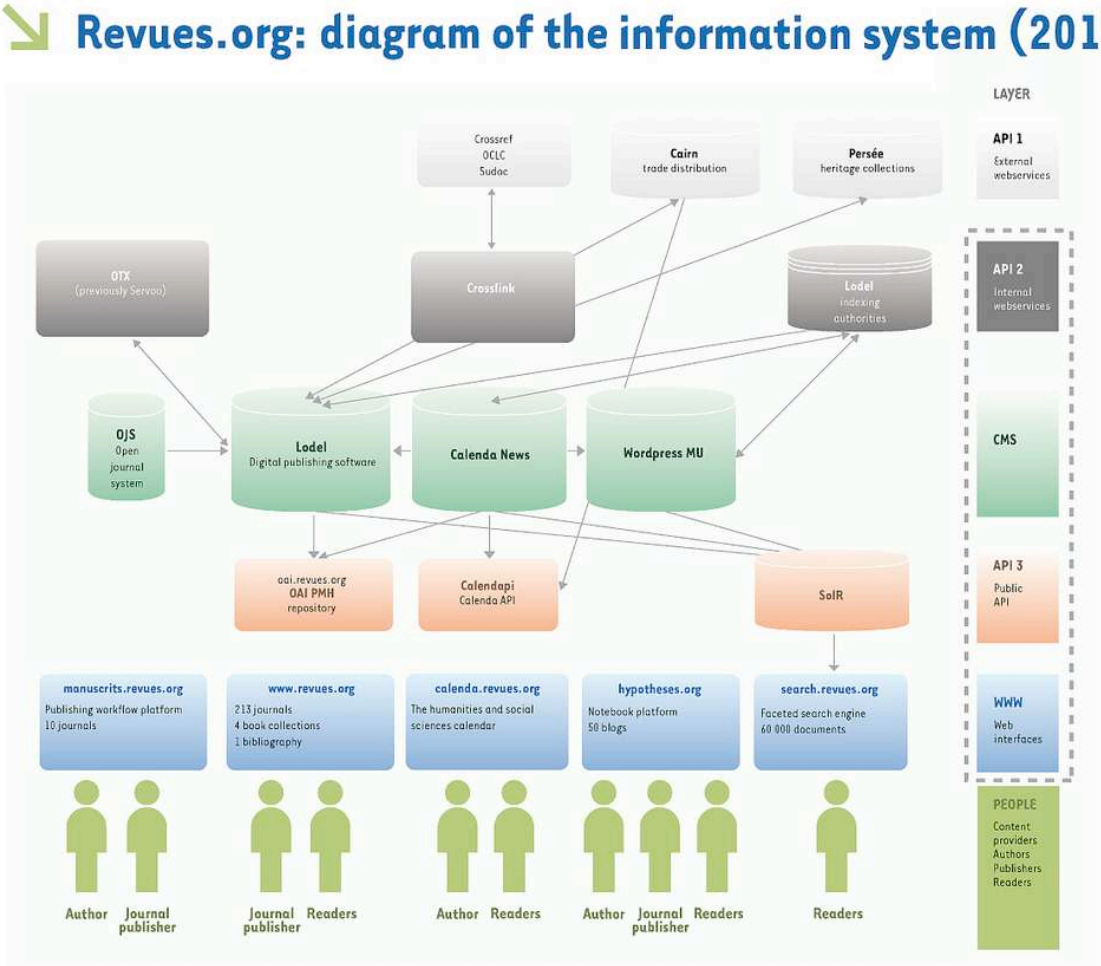
25 http://www.jisc.ac.uk/publications.aspx

concernés. Le premier exemple est un tableau présentant quelques standards et stratégies utilisés dans une cyberinfrastructure. Le second est le schéma simplifié du système d'information du Centre pour l'édition électronique, qui représente une toute petite brique de la cyberinfrastructure française en cours de constitution. Ce schéma laisse deviner la complexité de la cyberinfrastructure générale et permet d'illustrer la modularité de tout système d'information moderne.

Tableau de quelques technologies utilisées

Technologie	Finalité
Dublin Core	Description de métadonnées.
XML	Encodage de tous types de documents.
TEI, EAD, METS	Formats XML de référence.
RDF	Description des données (Web sémantique).
UNICODE	Jeu de caractère universel.
OAI-PMH	Protocole d'interopérabilité.
OPDS	Format d'échange de catalogues d'ouvrages.
DOI, HANDLE	Systèmes d'identification unique des documents.
OAIS	Principes d'archivage pérenne.
LOAD BALANCING	Système de répartition de charge pour assumer la progression forte d'un système.

Schéma simplifié du système d'information du Centre pour l'édition électronique ouverte



URBANISATION DU SYSTÈME D'INFORMATIONS

Une cyberinfrastructure doit faire appel au modèle d'urbanisation des systèmes d'information. Celui-ci a vocation à éviter la construction d'un dispositif monolithique, très peu évolutif et fragile. Au contraire, on privilégiera une architecture des données intégrant le principe de cohérence forte et de couplage faible. Il s'agit de prendre en compte l'ensemble du système d'information (SI) comme une ville qu'il faut lotir de façon optimale, considérant que la cohérence de l'ensemble doit être forte mais la dépendance entre les objets faible. Cela signifie qu'il faut monter des quartiers et des lotissements interdépendants mais assez autonomes pour pouvoir être remplacés individuellement sans mettre en péril l'équilibre de l'ensemble. Ce type de système n'est, bien entendu, pas cantonné à un périmètre fixé à l'avance. Il peut s'étendre et se reconfigurer en fonction de l'évolution des usages, des besoins et des environnements technologiques [26]. Les briques du dispositif communiquent entre elles comme avec les dispositifs extérieurs au système d'information : elles utilisent des API ou Webservices[27].

RISQUES ET DIFFICULTÉS

On aurait tort, cependant, de réduire la notion de cyberinfrastructure à un jeu de processus et de bonnes pratiques technologiques. Nous reprendrons le fil du Manifeste des Digital Humanities pour en énoncer les axes. Le premier risque concerne les difficultés d'accès aux données et aux métadonnées. En dehors des obstacles légitimes, notamment liés au respect de la législation sur les données nominatives, on peut craindre la mise en place de barrières institutionnelles ou commerciales. Mais d'autres freins existent, comme la faible documentation des données (article 9). Le deuxième risque porte sur la mise en place de secrets concernant les méthodes, le code et les formats qui ont abouti à la recherche (article 10). Même dans le milieu universitaire, et au sein d'équipes favorables au logiciel libre, il n'est pas si fréquent que du temps et de l'argent soient mobilisés pour la diffusion et la documentation de code[28]. Cela peut être le fait de négligences, d'un manque de moyens ou d'une volonté de protéger un pré carré. Le troisième obstacle porte sur la faible intégration des compétences liées aux humanités numériques dans les cursus universitaires et dans les établissements de recherche (article 11). Cela mènerait à un pilotage purement technologique, par des acteurs ignorant l'essentiel de ce qui constitue les disciplines au service desquelles est élaborée l'infrastructure. Sans appropriation par la profession dans son ensemble, la cyberinfrastructure sera plaquée sur des paradigmes et des concepts étrangers. On peut même craindre un dialogue de sourds entre ingénieurs et enseignants-chercheurs, ce qui constituerait une catastrophe. A l'inverse, nous avons besoin de médiateurs et de ponts

---

26 Akoka, Jacky, Isabelle Comyn-Wattiau, *et alii*, *Encyclopédie de l'informatique et des systèmes d'information*, Vuibert, Paris, 2006, 1941 p.

27 « Un service web est un programme informatique permettant la communication et l'échange de données entre applications et systèmes hétérogènes dans des environnements distribués. Il s'agit donc d'un ensemble de fonctionnalités exposées sur internet ou sur un intranet, par et pour des applications ou machines, sans intervention humaine, et en temps réel ». http://fr.wikipedia.org/wiki/Webservice Consulté le 7 janvier 2010.

28 Par exemple, sur SourceSup. http://sourcesup.cru.fr/

permettant de croiser, voire de faire converger, les dimensions de l'ingénierie et celles de la recherche. Cette question trouve des prolongements dans des secteurs qui ne sont pas discutés dans les conseils scientifiques des universités… alors qu'ils le devraient. Ils sont trop longtemps identifiés comme purement techniques, et se situent donc dans l'angle mort des débats au sein des établissements. Les conséquences peuvent en être fâcheuses. L'exemple du Digital Object Identifier (DOI) en est sans doute l'expression ultime, puisque la société Crossref (officiellement à but non lucratif) a réussi à en faire un bastion lui permettant d'imposer des règles de fonctionnement non scientifiques sur l'identification unique des documents scientifiques. Il existe des alternatives, mais celles-ci ne sont que faiblement soutenues et aucune organisation de l'ampleur de Crossref n'est, aujourd'hui, capable de contester son hégémonie. Or, Crossref impose une économie de la rareté à un bien informationnel stratégique, qui devrait être soumis à des principes scientifiques avant tout. On touche là aux questions de gouvernance du numérique, largement laissées à l'abandon par la communauté des historiens. Or, ce quatrième obstacle est sans doute le plus important (articles 13 et 14). L'histoire des rapports entre l'Etat et l'informatique est jalonnée d'exemples montrant de fortes difficultés d'adaptation à une situation très évolutive. De même, on ne devra pas perdre de vue que la plupart des grandes réussites de l'histoire de l'informatique n'avaient pas été prédites, ni anticipées. On proposera donc une approche incrémentale à un plan prédéfinit pour les vingt prochaines années sur une feuille blanche. Cette approche incrémentale devra répondre à des besoins réels, et ajuster le dispositif en permanence en fonction de l'évolution de ceux-ci. Il ne s'agit pas de mener une course sans fin à la nouveauté technologique, qui correspondrait à une gadgétisation du numérique. En revanche, la cyberinfrastructure devra s'attacher à la mise en place d'un *continuum* entre elle et les besoins de la recherche et de la société.

Conclusion

Le chantier qui s'ouvre durera dix, vingt ou trente ans. La route sera longue et complexe. L'ampleur des moyens à engager et la nécessité d'une vision à long terme ne doit, cependant, pas faire oublier que la science est faite d'imprévus et de retournements, elle doit donc rester essentiellement humaine, et permettre un séminaire permanent, selon le mot d'André Gunthert[29], et pas seulement ses concentrer sur les objets de la recherche qui sont les corpus et les produits de la recherche que sont les publications. Reléguer la conversation scientifique dans le marges des cyberinfrastructures serait une erreur stratégique majeure.

[29] **André** Gunthert, « Why Blog ? », in Marin Dacos (dir.), *Read/Write Book*, Marseille, Cléo (« Coll. Edition électronique »), 2010, [En ligne], mis en ligne le 25 mars 2010, Consulté le 23 janvier 2011. URL : http://cleo.revues.org/174